

数理统计与管理

Journal of Applied Statistics and Management

ISSN 1002-1566, CN 11-2242/01

《数理统计与管理》网络首发论文

题目: 双向聚类方法综述
作者: 方匡南, 陈远星, 张庆昭, 马双鸽
DOI: 10.13860/j.cnki.sltj.20190806-003
网络首发日期: 2019-08-19
引用格式: 方匡南, 陈远星, 张庆昭, 马双鸽. 双向聚类方法综述[J/OL]. 数理统计与管理. <https://doi.org/10.13860/j.cnki.sltj.20190806-003>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.13860/j.cnki.sljt.20190806-003

双向聚类方法综述

方匡南^{1,3} 陈远星¹ 张庆昭¹ 马双鸽^{1,2}

(1.厦门大学经济学院统计系,厦门福建 361005; 2.耶鲁大学生物统计系,纽黑文,美国, 06510; 3.新能源与储能运行控制国家重点实验室(中国电力科学研究院有限公司),北京 100192)

摘要:传统的聚类方法由于无法提取样本和变量间的局部对应关系,并且当数据具有高维性和稀疏性时表现不佳,因此学者们提出了双向聚类,基于样本和变量间的局部关系,同时对样本和变量进行聚类,形成一个子矩阵的聚类结果。近年来,双向聚类发展迅速,在基因分析、文本聚类、推荐系统等领域应用广泛。首先,对双向聚类方法进行梳理与归纳,重点阐述稀疏双向聚类、谱双向聚类和信息双向聚类三类方法,分析它们之间的区别和联系,并且介绍这三类方法在多源数据的整合分析、多层聚类、半监督学习以及集成学习上的发展现状和趋势;其次,重点介绍双向聚类在基因分析、文本聚类、推荐系统等领域的应用研究情况;最后,结合大数据时代的数据特征和双向聚类存在的问题,展望双向聚类未来的研究方向。

关键词: 稀疏双向聚类; 子矩阵; 谱双向聚类; 信息双向聚类

Review of Biclustering

FANG Kuang-nan^{1,3} CHEN Yuan-xing¹ ZHANG Qing-zhao¹ MA Shuang-ge^{1,2}

(1.Department of Statistics, School of Economics, Xiamen University, Xiamen Fujian, 361005, China; 2.Department of biostatistics, Yale University, New Haven, 06510,USA; 3.National Key Laboratory on Power Grid Environment Protection (China Electric Power Research Institute), Beijing 100192, China)

Abstract: The duality between samples and variables cannot be exploited by traditional clustering methods, which often performs badly when dealing with sparse and high-dimensional data. The scholars then proposed biclustering, which can cluster both samples and variables simultaneously to get submatrices, based on the duality between samples and variables. In recent years, biclustering methods have developed rapidly and been widely used in many areas, such as microarray analysis, text clustering, recommendation system and so on. This paper at first reviews biclustering methods and focus on three classical biclustering methods, including Sparse Biclustering, Spectral Biclustering and Information Theoretic Co-clustering. In detail, we conclude the differences and relationships between these three methods, and introduce the development status and trends of integrative analysis in multi-source datasets, multi-level clustering, semi-supervised learning, supervised learning and ensemble learning. Secondly, we focus on the application research of biclustering in the fields of microarray analysis, text clustering and recommendation system. Finally, combining the data characteristics of the big data era with the existing problems in biclustering, the future research direction of biclustering is discussed.

基金项目: 本文得到中央高校基本科研业务费专项资金资助(20720181003、20720171095)、新能源与储能运行控制国家重点实验室开放基金资助(NYB51201801579)。

Key words: sparse biclustering, submatrix, spectral biclustering, information theoretic co-clustering

0 引言

聚类分析(Clustering Analysis)是一种重要的无监督学习方法,它通过某个特定标准(如距离)将一个数据集划分为若干个子集,每个子集称为一个“簇”(Cluster)或者类。聚类分析的本质思想是将相似的数据对象聚为一类,差异大的数据对象划分到不同的类,从而将数据集划分为一系列具有不同模式的类别,即“物以类聚,人以群分”。根据数据对象的不同,传统的聚类方法可以分为两类:一类是基于所有变量对样本进行聚类(例如根据一些经济指标,将全国的城市分为一线、二线、三线城市),另一类是基于所有样本对变量进行聚类(例如根据肺癌患者的基因表达数据,将基因分为促进表达、抑制表达以及无影响三类),这两类方法仅对样本或变量聚类,因此常被称为单向聚类。单向聚类在对样本进行聚类时,一般基于所有变量来度量样本间的相似性,然而在某些情况下该方法存在一些局限性,例如在基因分析中,为了将肺癌病人细分为不同的子类,单向聚类基于所有基因(变量)对肺癌病人(样本)进行聚类,但实际中只有部分基因对肺癌有调控作用,并且这些具有调控作用的基因中,不同基因构成的基因组具体调控的肺癌子类也存在差异。因此需要在考虑不同基因组对肺癌子类的不同影响下进行聚类,即同时考虑病人和基因之间的局部对应关系。为了研究样本和变量之间的这种局部对应关系,学者们提出了双向聚类(Biclustering)。

双向聚类可以同时对本和变量进行聚类,将原始数据矩阵聚成一系列子矩阵,一个子矩阵表示一类。双向聚类的基本思想是研究样本和变量间的局部对应关系,通过确定其具体的结构来提高聚类效果并且增强聚类结果的可解释性。为了进一步解释双向聚类与传统单向聚类的区别,以基因数据为例,图 1-(a)表示肺癌基因矩阵,其中行和列分别表示样本和基因(变量),该矩阵中有 3 种阴影形状,每一种阴影形状表示一个基因组对肺癌子类的一种调控机制,非阴影区表示这些基因对肺癌没有影响。从图 1-(a)可以看出只有基因 A、D、I 对所有肺癌类型产生影响,倘若要基于所有的肺癌病人对基因进行聚类,那么结果将会如图 1-(b)所示,基因 A、D、I 聚为一类,表示这三个基因对所有肺癌病人均产生影响,这就是传统聚类的结果,该结果默认影响所有肺癌类型的基因组是相同的。然而,实际中影响不同肺癌类型的基因组存在差异,例如从图 1-(a)可以看出,基因 A、D、F、I 在样本 3、4 显示出一类调控机制,基因 A、D、I 在样本 1、5、8 显示出另一类调控机制,因此双向聚类将会产生例如图 1-(c)所示的部分聚类结果(由于篇幅限制,空白子矩阵省略)。

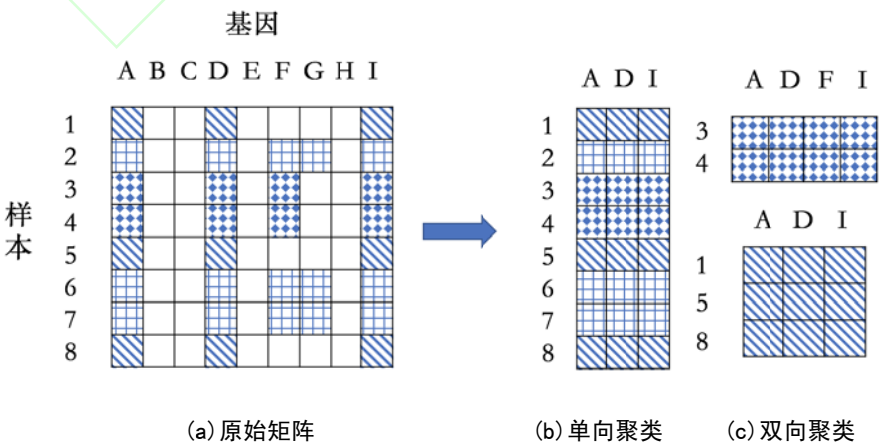


图 1 双向聚类与单向聚类示意图

双向聚类的目的是形成一系列子矩阵,根据该子矩阵是否存在具体结构分为两大类:结构矩阵和非结构矩阵。结构矩阵指聚类结果的子矩阵反映了一种潜在结构,例如常数值类、行常数(列常数)类、协同效应类等。非结构矩阵指可以用单向聚类中度量相似的指标来度量双向聚类下的相似性,此时的子矩阵并没有揭示隐藏的块结构,而是仅仅反映一种局部的相似性。

近年来双向聚类被广泛应用于文本挖掘、推荐系统、基因分析等诸多领域。在文本挖掘中,人们每时每刻都在互联网上传大量的文档,为了便于信息检索需要对这些文档进行有效地归类。一方面,这些文档构成的词频矩阵高维且稀疏;另一方面,一类文档往往在部分词语上表现相似。由于存在这两个特性,因此基于所有单词度量文档之间的相似性往往导致所有文档相似性都较低,进而导致对样本聚类效果不佳。而双向聚类通过寻找部分词语与部分文档的局部对应关系,可以快速有效地聚出相似的文档,并且通过其对应的高词频词汇来进行文档主题预测。在推荐系统中,大量用户在大量项目(例如商品)上的评分构成了高维稀疏的评分矩阵,比如淘宝、京东等电商为了精准地进行个性化推荐,就需要基于该评分矩阵对缺乏用户评价的商品进行评分预测,双向聚类通过考虑用户在某些商品上的评价相似性来进行预测,从而提高推荐效果。

本文的目的是系统梳理双向聚类方法的发展脉络、现状和趋势,重点阐述稀疏双向聚类、谱双向聚类和信息双向聚类三类方法,分析它们之间的区别和联系,介绍双向聚类在基因分析、文本聚类、推荐系统等领域的应用研究情况,并结合大数据时代的数据特征和双向聚类存在的问题,展望双向聚类未来的研究方向。

本文剩余部分安排如下:第二部分介绍稀疏双向聚类、谱双向聚类以及信息双向聚类三类经典的双向聚类方法,并且比较这三类方法的特点以及指出在多源数据的整合分析、多层聚类、半监督和监督学习以及集成学习上的发展;第三部分介绍这三类经典的双向聚类方法在基因分析、文本聚类、推荐系统等领域的应用情况;第四部分是总结全文并展望双向聚类的未来研究方向。

1 双向聚类方法

Hartigan 等(1972)^[1]最早研究双向聚类方法,提出“块聚类”(Block Clustering),该方法主要探究具有常数结构的子矩阵。而在基因数据中,存在一种子矩阵结构,反映某些样本在某些基因上存在一致上调或者一致下降的协同效应,因而 Cheng 和 Church(2000)^[2]提出了应用于高维基因数据的双向聚类方法。此后,学者们提出了许多双向聚类方法,包括结构和非结构双向聚类方法。由于篇幅限制,本文主要介绍三类经典的双向聚类方法:稀疏双向聚类(Sparse Biclustering, SB)、谱双向聚类(Spectral Biclustering, SBC)和信息双向聚类(Information Theoretic Co-clustering, ITCC)。

1.1 稀疏双向聚类

数据矩阵一般存在噪音,在噪音下寻找具有严格结构的子矩阵一般有两种方法:一种是设置一个阈值,通过衡量某个子矩阵的特定测量指标是否小于该阈值来确定该子矩阵是否符合特定结构;另一种则是构建一个具有严格结构的矩阵来近似原始矩阵,通过引入惩罚项来过滤噪音,从而形成近似矩阵,这种方法也称为稀疏双向聚类。它的目标函数基本形式如(1)式:

$$\operatorname{argmin} Loss + \lambda \times Penalty. \quad (1)$$

其中 $Loss$ 表示损失函数,一般用原始矩阵和近似矩阵之间的差异来刻画,我们希望损失函数尽量小; $Penalty$ 表示惩罚项,通过对损失函数中的参数施加惩罚从而将部分参数稀疏到0。目标函数整体反映损失和惩罚之间的权衡取舍,调节参数 λ 越大则惩罚力度越大。

稀疏双向聚类主要用来提取子矩阵的特定结构,其中包括常数值类、行(列)常数类和协同效应类(累加模型和累乘模型),详见图2。图2-(a)表示一个常数值类的子矩阵,在该子矩阵中所有元素都是常数1;图2-(b)表示一个行常数类的子矩阵,在该子矩阵中每一行的元素都是常数,不同行的元素可以不同,例如第一行元素都是1,第二行元素都是2,以此类推;图2-(c)表示一个协同效应类中的累加模型,在该子矩阵中任意一行的元素均是其他任意一行的元素同时加上一个数,任意一列的元素均是其他任意一列的元素同时加上一个数,例如第二行元素是第一行元素同时加上1,第二列元素是第一列元素同时加上1;图2-(d)表示一个协同效应类中的累乘模型,在该子矩阵中任意一行的元素均是其他任意一行元素的倍数,任意一列的元素均是其他任意一列元素的倍数,例如第二行元素是第一行元素的2倍,第二列元素是第一列元素的2倍。

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

(a) 常数值类

1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4

(b) 行常数类

1	2	4	5
2	3	5	6
4	5	7	8
8	9	11	12

(c) 协同效应类
-累加模型

1	2	3	4
2	4	6	8
4	8	12	16
6	12	18	24

(d) 协同效应类
-累乘模型

图2 双向聚类结构

1.1.1 稀疏奇异值分解

Lee等(2010)^[3]提出了基于稀疏奇异值分解进行双向聚类。给定基因矩阵 $X_{n \times d}$,则矩阵 X 的奇异值分解如式(2):

$$X = UDV' = \sum_{t=1}^m s_t u_t v_t' \quad (2)$$

其中 m 表示 X 的秩, $U = (u_1, \dots, u_m)$ 表示左奇异向量构成的矩阵, $V = (v_1, \dots, v_m)$ 表示右奇异向量构成的矩阵, $D = \text{diag}(s_1, \dots, s_m)$ 是一个对角矩阵,对角线上的元素为非负的奇异值 $s_1 \geq \dots \geq s_m$ 。实际上,奇异值分解将一个矩阵分解为 m 个秩为1的矩阵 $s_t u_t v_t'$ 之和,由于奇异值分解可以分离出主要信息,取值越大的奇异值对应的秩为1的矩阵含有的信息越多,因此可以考虑提取奇异值最大的前 T 个矩阵,又称为前 T 层,如式(3):

$$X \approx X^{(T)} = \sum_{t=1}^T s_t u_t v_t' \quad (3)$$

根据奇异值从大到小逐层进行子矩阵提取,限于篇幅,以提取第一层为例介绍其算法,如需提取多层,可将矩阵 X 减去提取的第一层矩阵作为提取第二层的基础矩阵,以此类推提取后续的层。事实上,若是不施加惩罚地提取第一层,则对应奇异值最大的矩阵并不具有稀疏性,无法形成目标子矩阵,因而考虑对左右奇异向量施加惩罚,从而使得向量中不重要的元素稀疏到0,最终左奇异向量(右奇异向量)中非0元素锁定聚类子矩阵中的样本(变量)。目标函数如式(4):

$$\arg \min_{s, u, v} \|X - suv'\|_F^2 + \lambda_u P_1(su) + \lambda_v P_2(sv) \quad (4)$$

其中 $\|\cdot\|_F$ 表示Frobenius范数, s, u, v 分别表示第一层奇异值、左奇异向量和右奇异向量, $P_1(\cdot)$ 和 $P_2(\cdot)$ 分别表示对 u 和 v 的惩罚, λ_u 和 λ_v 为对应的非负调节参数。

基于上述稀疏奇异值分解的框架, Sill 等(2011)^[4]结合稳健性选择来进行参数调节, 从而可以在控制第一类错误概率的前提下进行双向聚类。Lee 和 Huang(2014)^[5]则是将稀疏奇异值分解应用到二元矩阵的双向聚类上。Sun 等(2014)^[6]提出多视角双向聚类(Multi-view Biclustering), 将稀疏奇异值分解拓展到多个数据集, 通过锁定相同样本探究基因和临床表现的关系。Sun 等(2015)^[7]指出 Sun 等(2014)^[6]的算法无法保证收敛, 因而提出一个全局收敛算法, 并且将 Lasso 惩罚改成 L_0 惩罚, 得到锁定相同样本下的两个子矩阵。Min 等(2016)^[8]为了探究 miRNA 和 mRNA 之间的关系, 采用两步法, 第一步通过对病人测量的 miRNA 和 mRNA 数据集进行多元稀疏组惩罚来得到 miRNA 和 mRNA 之间的关系矩阵, 第二步对该矩阵进行 L_0 的稀疏奇异值分解, 得到最终的结果。

1.1.2 稀疏双向 K-Means 聚类

Tan 和 Witten(2014)^[9]借鉴稀疏 K-Means 聚类^[10]的想法, 将其拓展到双向聚类中。给定数据矩阵 $X_{n \times d}$, 假设 n 个样本属于 K 个互不交叉的类(C_1, \dots, C_K), d 个变量属于 R 个互不交叉的类(D_1, \dots, D_R), 其中 $C_k \subset \{1, \dots, n\}$ 且 $D_r \subset \{1, \dots, d\}$, 分别表示样本和变量的类别标签。同时假设矩阵中的元素互相独立且 $X_{ij} \sim N(\mu_{kr}, \sigma^2)$, μ_{kr} 表示第 k 个样本类和第 r 个变量类所对应的子矩阵中所有元素的均值, 最终的聚类目标是估计 C_k 、 D_r 以及 μ_{kr} , 目标函数如式(5):

$$\arg \min \{ \sum_{k=1}^K \sum_{r=1}^R \sum_{i \in C_k} \sum_{j \in D_r} (X_{ij} - \mu_{kr})^2 \}, \quad (5)$$

当 $R = 1$, 该问题转化为样本 K-Means 聚类; 当 $K = 1$, 该问题转化为变量 K-Means 聚类。求解目标函数(5)最终会形成 KR 个类式, 并且每个类的均值 $\mu_{kr} \neq 0$ 。现实中部分类的实际均值等于 X 的总体均值 μ , 而在噪音的影响下表现为接近 μ 的均值, 而聚类的目标是探究突出于平均水平的类别(即明显高于或者低于总体均值的类), 而非那些居于总体平均水平的类。换言之, 假设对矩阵 X 进行中心化, 使得矩阵中所有元素的总体均值为 0, 则可能存在一些类别的均值接近 0, 为了增强聚类结果的可解释性并且减少类别方差, 需要将这些均值接近 0 的类别均值压缩到 0, 因此得目标函数:

$$\arg \min \{ \frac{1}{2} \sum_{k=1}^K \sum_{r=1}^R \sum_{i \in C_k} \sum_{j \in D_r} (X_{ij} - \mu_{kr})^2 + \lambda \sum_{k=1}^K \sum_{r=1}^R |\mu_{kr}| \}. \quad (6)$$

其中 λ 是一个非负的调节参数, 通过对式(5)中的均值 μ_{kr} 施加 Lasso 惩罚从而使不重要的类均值稀疏到 0, 因而最终的聚类结果仅仅呈现那些均值明显偏离 0 的类。

类似 K-Means 聚类的算法, 稀疏双向 K-Means 聚类可以利用迭代算法求解得到局部最优解, 只是在执行算法前需要对矩阵 X 中心化, 得到矩阵 X^* , 再对 X^* 执行算法。

由于 Tan 和 Witten(2014)^[9]提出的稀疏双向 K-Means 聚类需要对矩阵元素进行正态假设来寻找常数子矩阵, Pimentel 等(2018)^[11]将稀疏性典型相关分析(Sparse canonical correlation analysis; SCCA)应用到双向聚类上, 无需进行正态假定, 仅利用样本和基因的局部相关性即可聚出多种类型的子矩阵。由于稀疏双向 K-Means 聚类产生非交叉的聚类结果(即每个样本和基因只属于一个类), Teng 等(2016)^[12]在其基础上考虑产生交叉的聚类结果(每个样本和基因允许同时归属于多个类)。

1.1.3 凸双向聚类

稀疏奇异值分解和稀疏双向 K-Means 聚类算法只能得到局部最优解, 聚类结果会随初始值的变化而变化。为了解决该问题, Chi 等(2016)^[13]基于凸目标函数和 Fused Lasso 惩罚函数提出了凸双向聚类(Convex Biclustering)。此方法基于凸规划可以得到全局最优解, 因此聚类结果不受初值影响, 同时随着调节参数的变化可以形成一条类似于层次聚类的聚类路径, 便于观察聚类结果的变化。具体而言, 凸双向聚类试图拟合一个矩阵 U 来近似数据矩阵 X , 这与稀疏奇异值分解类似。不同的是, 稀疏奇异值分解是利用 SVD 分解的结果施加惩罚来达

到稀疏的效果，而这里则是同时对不同行向量的差值和不同列向量的差值施加惩罚，从而使得相似的行合并成类，以及相似的列合并成类。如果仅考虑行或列的差值，则该问题转化为凸聚类问题^[14]。凸双向聚类同时考虑行和列的差值相当于综合考虑行和列的联合影响，因此最终形成 KR 个常数子矩阵。具体目标函数：

$$F_{\gamma}(U) = \frac{1}{2} \|X - U\|_F^2 + \gamma[\Omega_w(U) + \Omega_w(U')]. \quad (7)$$

其中 γ 表示非负的调节参数， $\Omega_w(U) = \sum_{i < j} w_{ij} \|U_{\cdot i} - U_{\cdot j}\|_2$ ， w_{ij} 表示权重且 $U_{\cdot i}(U_i)$ 表示矩阵 U 第 i 列(行)。基于惩罚项的形式可知，该目标函数促进类似的行和列趋于相同，从而形成常数子矩阵。 w_{ij} 使用稀疏高斯核权重(Sparse Gaussian Kernel Weights)提升运行效率，并且利用 DLPA 算法将凸双向聚类问题转化为求解单向的凸聚类问题，再用 AMA 算法分别求解行和列的凸聚类，不断迭代得到最终结果。

Nezhad 等(2017)^[15]将凸双向聚类算法拓展到监督学习的精准医疗中。Chi 等(2018)^[16]将凸双向聚类算法推广到张量数据上。Yu 等(2018)^[17]将凸双向聚类算法推广到多任务学习中，一方面对多个因变量和多个自变量之间的关系进行参数估计，另一方面对参数矩阵进行双向聚类探究自变量和因变量之间的群组关系。

1.2 谱双向聚类

谱聚类(Spectral Clustering)是建立在图论的基础上，考虑图形最优切割问题而引申出来的聚类方法。相比于K-Means只能对线性可分的类别形状进行聚类，谱聚类可以对非线性可分的类别形状进行有效聚类。以图3-(a)为例，每个方形代表一个样本，图中有6个样本，样本之间以边连接，每条边赋予权重。权重一般是根据样本之间的距离大小确定，距离越小权重越大，即样本越相似。谱聚类的最终目的是确定类似图中虚线的分割线，从而将不同类别的点分割开。例如图3-(a)中，这条虚线将样本分为2类， a_2 和 a_3 归为一类，其余4个样本归为另一类。关于谱聚类的具体细节详见Luxburg (2007)^[18]

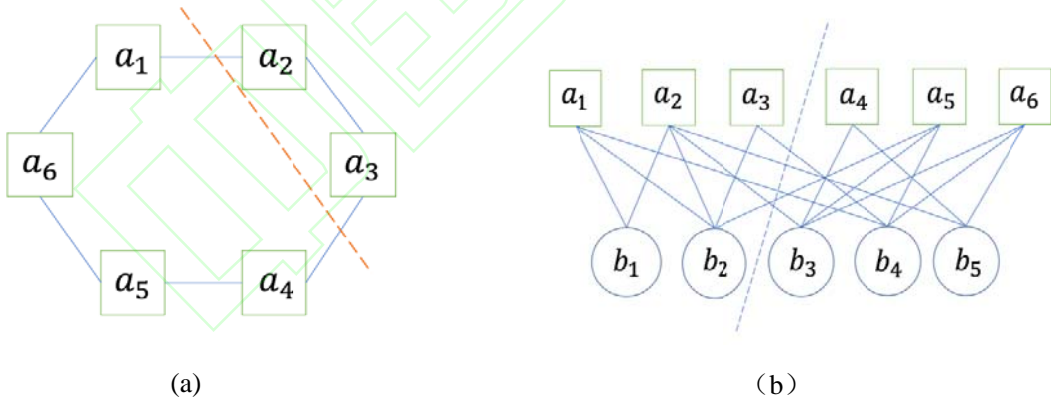


图3 谱聚类 and 双向谱聚类示意图

基于谱聚类的基本思想，Dhillon (2001)^[19]将其拓展到双向聚类中，谱聚类基于样本间的距离建立邻接矩阵度量样本相似性，而双向谱聚类则是基于样本和变量间的关联性建立邻接矩阵。考虑矩阵 $X_{n \times d}$ 构建二分图(bipartite graph) G ，其中 $\{a_1, \dots, a_n\}$ 表示行节点， $\{b_1, \dots, b_d\}$ 表示列节点，行节点和列节点之间以边连线(边权重对应 X_{ij})，行节点之间以及列节点之间无法连线。二分图如图3-(b)所示，方形代表样本(行节点)，圆形代表变量(列节点)，图中有6个样本，5个变量，只有样本和变量之间以边连接，因此谱双向聚类的目的是寻找一个分割。如图3-(b)所示，样本 a_1 、 a_2 、 a_3 和变量 b_1 、 b_2 聚为一类，其余的样本和变量聚为另一类，因此每一类对应的部分样本和部分变量形成一个子矩阵，该聚类结果正是基于样本和变量的关联性(以 X_{ij} 体现)。

类比谱聚类构建拉普拉斯矩阵的过程，谱双向聚类构造两个对角矩阵 H_1 和 H_2 ，矩阵 H_1 中第 i 个对角元素为 $\sum_{j=1}^d X_{ij}$ ，矩阵 H_2 中第 j 个对角元素为 $\sum_{i=1}^n X_{ij}$ ，因此形成如下的程度矩阵 (degree matrix) 和邻接矩阵 (adjacent matrix)：

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \quad W = \begin{pmatrix} 0 & X \\ X' & 0 \end{pmatrix}$$

从而得到拉普拉斯矩阵：

$$L = H - W$$

当使用标准化分割目标函数 (Normalized-cut Objectives) 时，可以证明，二分类时目标函数的解为特征问题： $Lz = \lambda Hz$ 的解中第二小的特征值对应的特征向量。此时最小化分割问题转化为求解 $(n + d) \times (n + d)$ 矩阵 $H^{-1}L$ 的特征值和特征向量。基于二分图的结构，文中将目标函数转化为奇异值分解的问题，解得 $X_{new} = H_1^{-1/2} X H_2^{-1/2}$ 的第二大奇异值对应的左右奇异向量 u_2 和 v_2 ，从而得到最优解为向量 $z = \begin{pmatrix} H_1^{-1/2} u_2 \\ H_2^{-1/2} v_2 \end{pmatrix}$ 。鉴于矩阵 X_{new} 为 $n \times d$ 维矩阵，因此利用奇异值分解有效降低矩阵维度从而减少计算量。在得到解向量后，通过该向量进行 K-Means 聚类得到最终的 2 个类别。同理，当分类个数 $l > 2$ 时，可以对矩阵进行 SVD 分解确定除最大奇异值外的 $\lceil \log_2 l \rceil$ 个最大奇异值对应的左右奇异向量 U 和 V ($\lceil \cdot \rceil$ 表示取整)。之后基于左右奇异向量构建矩阵 $z = \begin{pmatrix} H_1^{-1/2} U \\ H_2^{-1/2} V \end{pmatrix}$ ，并且基于该矩阵进行 K-Means 聚类，得到 l 个类。最终聚类结果显示部分样本和部分变量共属一类，因而形成矩阵块的聚类结构，即完成了双向聚类。

在 Dhillon (2001)^[19] 提出双向谱聚类之后，许多学者对该方法进行拓展。针对 K-Means 聚类依赖于初值导致结果不稳定的缺点，Nie 等 (2017)^[20] 对邻接矩阵寻找严格块对角的近似矩阵，因此基于该近似矩阵构造的拉普拉斯矩阵进行特征分解可以直接得到聚类结果而无需 K-Means 聚类。在探索三阶数据 (例如用户、搜索内容、搜索网页三个层次的数据) 的相似类别时，Gao 等 (2005)^[21] 基于谱双向聚类的原理，将其拓展到三阶数据的情形，即考虑同时对三个层次进行聚类。在处理动态演变的数据时，Green 等 (2011)^[22] 考虑对随时间推移变化的矩阵进行谱双向聚类，基本思想是对不同时间点聚类结果进行平滑，同时保留现有时间点数据特有的结构。Shi 等 (2011)^[23] 考虑对于半监督的文本聚类，引入限制矩阵刻画已知的文本和词语的关系，形成新型拉普拉斯矩阵从而得到聚类结果。Huang 等 (2015)^[24] 考虑到当前许多双向聚类方法得到的结果参差不齐，借鉴聚类方法中的集成算法，将多个双向聚类算法的结果利用谱双向聚类进行集成得到最终相对一致的聚类结果。谱双向聚类还可以应用于基因分析中，Kluger 等 (2003)^[25] 考虑基因矩阵 $X_{n \times d}$ ，将每个基因在特定样本下的表达值分解为： $X_{ij} = E_{ij} \times \rho_i \times \chi_j$ ，通过谱双向聚类得到基因的协同效应矩阵。Cano 等 (2007)^[26] 在探究如何形成基因表达不足的子矩阵时，将改良概率聚类 (Improved Possibilistic Clustering) 和谱双向聚类进行结合，实现聚类结果可交叉的目的。除此之外，Guan 等 (2005)^[27]、Wieling 和 Nerbonne (2009)^[28] 以及 Xu 等 (2010)^[29] 将谱双向聚类分别应用到图像分类、方言种类区分以及网页浏览上。

1.3 信息双向聚类

Dhillon 等 (2003)^[30] 基于交互信息损失最小原则寻找一个对原始矩阵的最优近似，从而提出信息双向聚类。相对于谱双向聚类基于图形划分构建目标函数，使得聚类结果中变量类别与样本类别一一对应，信息双向聚类则是放宽该限制，允许样本类别与多个变量类别相对应，即允许出现交叉结构。在该方法下，数据矩阵的行和列可视为两个随机变量，于是矩阵

元素反映一个二元联合分布的概率，通过引入交互信息 (Mutual Information) 的概念，将原始矩阵中的行和列映射到行和列的类别变量上，利用损失的交互信息最小为目标构建目标函数。

具体而言，令 A 和 B 分别表示取值为 $\{a_1, \dots, a_n\}$ 和 $\{b_1, \dots, b_d\}$ 的随机变量，则 $p(A, B)$ 表示联合概率分布。假设将 A 分成 K 类， B 分成 R 类，令 $\{\hat{a}_1, \dots, \hat{a}_K\}$ 表示 A 的 K 个类，令 $\{\hat{b}_1, \dots, \hat{b}_R\}$ 表示 B 的 R 个类，则目的是找到一个映射 M_A 和 M_B ，使得

$$\hat{A} = M_A: \{a_1, \dots, a_n\} \rightarrow \{\hat{a}_1, \dots, \hat{a}_K\} \quad \hat{B} = M_B: \{b_1, \dots, b_d\} \rightarrow \{\hat{b}_1, \dots, \hat{b}_R\}$$

其中 (M_A, M_B) 表示双向聚类的结果。由于交互信息 $I(A, B)$ 反映变量 A 包含与 B 相关的信息量 (反之亦然)，因此聚类结果的优劣可以基于原始矩阵映射到新矩阵后交互信息损失大小来度量。通过证明，对于一个固定的双向聚类结果 (M_A, M_B) ，交互信息的损失可以写成：

$$I(A, B) - I(\hat{A}, \hat{B}) = O(p(A, B) || q(A, B)), \quad (8)$$

这里 $O(||)$ 表示 KL 散度 (相对熵)， $q(A, B)$ 的分布如式 (9)：

$$q(a, b) = p(\hat{a}, \hat{b})p(a|\hat{a})p(b|\hat{b}), \quad a \in \hat{a}, b \in \hat{b}. \quad (9)$$

式 (8) 和式 (9) 指出，综合考虑 A 和 B 的联合分布和边际分布可得出新的联合分布 $q(A, B)$ ，可以证明， $p(A, B)$ 和 $q(A, B)$ 的 KL 散度可以表示成 $p(B|a)$ 和 $p(B|\hat{a})$ 相对熵的加权平均，也可以表示成 $p(A|b)$ 和 $p(A|\hat{b})$ 相对熵的加权平均。因此可以通过给定一个方向的聚类结果，对另一个方向进行聚类，不断重复直至收敛。

信息双向聚类通过引入交互信息损失来构建近似矩阵，从而得到双向聚类的结果。基于交互信息损失最小的想法，Ghosh (2007)^[31] 提出一个基于布雷格曼散度 (Bregman divergence) 的广义双向聚类框架，该方法将近似误差用布雷格曼散度来衡量，由于布雷格曼散度包括了 KL 散度，因此统一了信息双向聚类的基本框架。Ye 等 (2009)^[32] 通过对变量进行加权，对重要变量赋予更大权重，从而提出变量加权信息双向聚类 (Feature Weighting Information-Theoretic Co-clustering; FWITCC)。在处理高于三个层次的数据时，Chiaravalloti 等 (2006)^[33] 考虑两两关系，并且将已有的信息作为约束，将问题转化半监督三向聚类进行优化求解。在探究如何高效地处理海量数据时，Huang 等 (2012)^[34] 将数据进行分块从而进行分布式计算，再将每一块数据的聚类结果进行集成从而得到良好的一致聚类结果。信息双向聚类还可以应用在其他领域，例如 Cai 等 (2004)^[35] 和 Wang 等 (2005)^[36] 分别将信息双向聚类应用到声音场景分类和视频截图的分类上，提高了这两个领域的分类效果。Ahmed 和 Mahmood (2014)^[37] 利用信息双向聚类来检测拒绝服务攻击 (DoS Attack)，从而提高检测效果。

1.4 双向聚类方法比较

表 1 对上述的三类双向聚类方法进行了总结和比较，稀疏双向聚类中的基于稀疏奇异值分解方法得到协同效应的子矩阵，而稀疏双向 K-Means 聚类和凸双向聚类得到常数子矩阵，这三种方法共同特征都是损失函数加上惩罚函数来得到稀疏聚类结果，并且由于存在具体的聚类结构，一般主要应用于对基因结构的分析中。谱双向聚类则是基于图论中的二分图，利用 NCut 最优切割来得到聚类结果，该聚类结果分为两类：非结构的子矩阵或协同效应子矩阵。前者一般应用于文本聚类、用户推荐上，由于这两个领域仅关注在部分变量下表现相似的样本，无需探究具体的子矩阵结构；后者一般应用于对基因结构的分析中。信息双向聚类则是基于信息论中的 KL 散度，利用交互信息损失最小来得到聚类结果，该聚类结果是非结构的子矩阵，一般应用于文本聚类和用户推荐上。关于这三类双向聚类方法的程序也在表 1 中列出，读者可以自行查看。

表 1 双向聚类方法的联系与区别

方法	参考文献	聚类结构	基本思想	软件
稀疏奇异值分解	[3-8]	协同效应	双向 Lasso 惩罚的奇异值分解	https://cran.r-project.org/web/packages/s4vd/index.html
稀疏双向 K-Means 聚类	[9] [11-12]	常数	双向 Lasso 惩罚的 K-Means 聚类	https://cran.r-project.org/web/packages/sparseBC/index.html
凸双向聚类	[13] [15-17]	常数	双向 Fused Lasso 的凸函数近似	https://cran.r-project.org/web/packages/cvxbiclustr/index.html
谱双向聚类	[25-26] [19-24] [27-29]	协同效应 无	基于二分图 Ncut 最优切割准则	基因聚类: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.bicluster.SpectralBiclustering.html 文本聚类: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.bicluster.SpectralCoclustering.html
信息双向聚类	[30-37]	无	基于信息论 KL 散度的交互信息损失最小准则	https://github.com/slongwell/ITCC

2 应用分析

双向聚类方法可以广泛应用于多个领域，但限于篇幅，本文主要以基因分析、文本聚类以及推荐系统三个领域为例介绍其实际应用情况。

在基因分析中，由于部分基因对不同疾病类型产生不同影响，并且基因矩阵是个高维矩阵，因此双向聚类被普遍应用于基因分析中。Lee 等 (2010)^[3]将稀疏奇异值分解应用于肺癌基因数据，对分解得到的前三层左奇异向量(对应样本)进行两两组合成坐标系，基于这三个维度有效地将不同的肺癌类型区分开，并且通过提取三层的组合热力图可以和实际的类别标签相匹配进一步验证其效果。Tan 和 Witten (2014)^[9]将稀疏双向 K-Means 聚类应用于肺癌基因数据(和 Lee 等 (2010)^[3]应用数据集一致)，聚类结果的热力图与稀疏奇异值分解提取前三层的热力图类似，利用稀疏奇异值分解得到稀疏向量，需要进行后续处理得到具体的类别，然而该方法可以直接得到类别标签。Chi 等 (2016)^[13]将凸双向聚类应用在同一个肺癌数据中，对数据集加上不同幅度噪音来检测该方法的稳健性。具体而言，文中对原始的肺癌数据矩阵加上独立同分布的 $N(0, \sigma^2)$ 噪音， σ 分别设置为 0.5, 1 和 1.5。兰德指数 (Rand Index) 显示凸双向聚类的稳定性和可重复性相比于前两种方法较好。Sun 等 (2014)^[6]将稀疏奇异值分解应用在两个数据集的整合分析上，两个数据集分别测量非裔美国人的基因数据和临床表征数据，最终聚出三组样本及其对应的基因和临床表征变量，利用前两组的变量进行逻辑回归，发现在预测是否具有可卡因依赖性上相比控制组的变量具有高的预测准确率。

在文本聚类中，由于文本矩阵是高维稀疏的矩阵，并且部分词语仅在部分文档中显著出现，从而利用这些高频词汇可以推测文本主题，因此双向聚类也被普遍应用于文本聚类中。

Dhillon (2001)^[19]将谱双向聚类用于“SMART”数据集的聚类, 通过将医学类、航空系统类以及信息提取类的文档进行两两组合、三者组合, 从而验证谱双向聚类在二分类及多分类问题中的高准确率。Dhillon 等 (2003)^[30]将信息双向聚类用于同一个数据集, 验证了信息双向聚类相比于单向聚类在分类准确率上的显著优势。Castro 等 (2007)^[38]基于人体免疫系统提出双向聚类方法, 并且利用从巴西报纸上摘取的文档作为数据集进行验证, 结果显示该方法比传统的 K-Means 聚类效果有所提升。Nie 等 (2017)^[20]将谱双向聚类进行拓展, 提出 SOBG 算法, 文中验证了该算法在 4 个不同数据集下相比谱双向聚类以及其他单向聚类准确率得到提升, 并且越是高维的数据提升效果越明显。

在推荐系统中, 由于推荐矩阵是高维稀疏的矩阵, 并且部分用户往往在某类项目上存在一致的评价, 因此双向聚类也应用于推荐系统中。Liang 和 Leng (2014)^[39]将信息双向聚类应用于用户推荐上, 通过两个电影数据集进行检验, 结果显示侧重类别偏好并且重点基于电影进行推荐的效果最好, 而基于用户和电影的双向信息聚类得到的预测效果, 要显著优于基于用户 (User-based) 和基于项目 (Item-based) 这两类方法的预测效果。Wang 等 (2015)^[40]提出 BC-CF 双向聚类方法, 利用 MovieLens 数据集进行检验, 结果显示该方法能够解决数据稀疏性的问题, 并且预测精度要高于传统的基于用户的预测方法。Alqadah 等 (2015)^[41]提出 BCN 的双向聚类方法, 该方法用来推荐前 N 个项目给用户, 在稀疏数据下 BCN 显著优于传统的基于项目的预测方法, 并且效果优于同样可以处理稀疏数据的 SLIM 和 WRMF 方法。

3 总结与讨论

双向聚类可以同时对本体和变量进行聚类, 将原始数据矩阵聚成一系列子矩阵, 一个子矩阵表示一类。双向聚类的基本思想是分析样本和变量间的局部对应关系, 通过确定其具体的结构来提高聚类效果并且增强聚类结果的可解释性。本文主要梳理了稀疏双向聚类、谱双向聚类以及信息双向聚类这三类经典方法。在这三类方法中, 稀疏双向聚类主要是基于损失函数+惩罚函数的框架, 在拟合优势和稀疏惩罚之间进行权衡取舍, 最终得到稀疏聚类结果。谱双向聚类通过将聚类问题和图论中的二分图切割相结合, 基于奇异值分解得到最优解。信息双向聚类基于信息论, 考虑将原始矩阵视为二元联合概率的结果并进行近似, 利用交互信息损失最小准则进行优化。这三类方法后来被推广到高阶多层次数据的聚类、多个数据集的整合分析以及半监督、监督学习和集成学习中。最后梳理了双向聚类在基因分析、文本聚类、用户推荐等领域的应用情况。

近些年随着大数据时代的到来, 实时更新的海量数据需要通过人工智能及时处理。例如不断更新上传文档的分类从而进行快速的信息检索, 不断更新用户对商品的评价从而进行个性化推荐等, 这些领域都需要利用双向聚类来不断地进行建模分析, 并且对结果进行跟踪反馈和更新调整。基于对现有研究问题的剖析和总结, 我们认为双向聚类未来可能的研究方向有:

1) 结构自动识别的双向聚类

现有的双向聚类方法一般事先给定需要聚类的子矩阵结构, 并且构建目标函数进行优化求解。然而现实中数据结构是复杂多样的, 我们事先并不能确定数据反映的模式结构, 直接给定结构可能导致聚类结果和实际情况相去甚远。因此有必要研究一种方法, 可以自动识别数据中的子矩阵结构并且定位到该子矩阵, 从而可以根据具体的数据结构自发地探究隐含的内在结构。

2) 多源数据融合的双向聚类

现有的双向聚类方法已经被应用于多个数据集的整合分析^[6-8], 这些方法中的多个数据集对应相同的样本, 不同的数据集变量类型不同, 通过锁定相同的样本来探究多个变量组之

间的关系。现实中搜集到的多个数据集往往在样本上也存在差异,因此有必要设计一种方法,使得锁定不同数据集中的相似的样本,进而探究多个变量组之间的关系。

3) 复杂数据的双向聚类

随着大数据的兴起,除了数据量的指数级递增以外,还伴随着数据结构复杂性的增加,现有的双向聚类方法一般研究静态的、低频的数据矩阵。然而函数型数据、高频数据、面板数据以及网络数据等复杂的数据结构已经逐渐成为统计研究的热点(王德青等, 2015^[42]; 王德青等, 2015^[43]; 林秀梅等, 2016^[44])。比如 Slimen 等(2018)^[45]已经开始研究对函数型数据进行双向聚类,将双向聚类应用在移动网络信号的模式聚类中。综上所述,如何将双向聚类方法拓展到适用于复杂的数据类型将成为双向聚类未来研究的一个重要方向。

参考文献

- [1] Hartigan J A. Direct Clustering of a Data Matrix[J]. Journal of the American Statistical Association, 1972, 67(337):123-129.
- [2] Cheng Y, Church G M. Biclustering of expression data[C]. Eighth International Conference on Intelligent Systems for Molecular Biology. AAAI, 2000:93-103.
- [3] Lee M, Shen H, Huang J Z, et al. Biclustering via sparse singular value decomposition[J]. Biometrics, 2010, 66(4):1087-1095.
- [4] Sill M, Kaiser S, Benner A, et al. Robust biclustering by sparse singular value decomposition incorporating stability selection[J]. Bioinformatics, 2011, 27(27):2089-2097.
- [5] Lee S, Huang J Z. A biclustering algorithm for binary matrices based on penalized Bernoulli likelihood[J]. Statistics & Computing, 2014, 24(3):429-441.
- [6] Sun J, Bi J, Kranzler H R. Multi-view singular value decomposition for disease subtyping and genetic associations[J]. BMC Genetics, 2014, 15(1):1-12.
- [7] Sun J, Lu J, Xu T, et al. Multi-view sparse co-clustering via proximal alternating linearized minimization[C]. Proceeding of the 32nd International Conference on Machine Learning. ICML, 2015:757-766.
- [8] Min W, Liu J, Luo F, et al. A Two-Stage Method to Identify Joint Modules From Matched MicroRNA and mRNA Expression Data[J]. IEEE Transactions on Nanobioscience, 2016, 15(4):362-370.
- [9] Tan K M, Witten D M. Sparse Biclustering of Transposable Data[J]. Journal of Computational & Graphical Statistics, 2014, 23(4):985-1008.
- [10] Witten D M, Tibshirani R. A Framework for Feature Selection in Clustering[J]. Journal of the American Statistical Association, 2010, 105(490):713-726.
- [11] Pimentel H, Hu Z, Huang H. Biclustering by sparse canonical correlation analysis[J]. Quantitative Biology, 2018, 6(1):1-12.
- [12] Teng B, Yang C, Liu J, et al. Exploring the Genetic Patterns of Complex Diseases via the Integrative Genome-Wide Approach[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2016, 13(3):557-564.
- [13] Chi E C, Allen G I, Baraniuk R G. Convex Biclustering[J]. Biometrics, 2016, 73(1):10-19.
- [14] Chi E C, Lange K. Splitting Methods for Convex Clustering[J]. Journal of Computational & Graphical Statistics.
- [15] Nezhad M Z, Zhu D, Sadati N, et al. SUBIC: A Supervised Bi-Clustering Approach for Precision Medicine[C]. 16th IEEE International Conference on Machine Learning and Applications. ICMLA, 2017.
- [16] Chi E C, Gaines B R, Sun W W, et al. Provable Convex Co-clustering of Tensors.arXiv preprint arXiv:1803.06518, 2018.

- [17] Yu M, Ramamurthy K N, Thompson A, et al. Simultaneous Parameter Learning and Bi-Clustering for Multi-Response Models. arXiv preprint arXiv:1804.10961, 2018.
- [18] Luxburg U V. A tutorial on spectral clustering[J]. Statistics & Computing, 2007, 17(4):395-416.
- [19] Dhillon I S. Co-clustering documents and words using bipartite spectral graph partitioning[C]. ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.ACM, 2001:269-274.
- [20] Nie F, Wang X, Deng C, et al. Learning A Structured Optimal Bipartite Graph for Co-Clustering[C].Advances in Neural Information Processing Systems. NIPS, 2017.
- [21] Gao B, Liu T Y, Zheng X, et al. Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering[C]. Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. ACM, 2005:41-50.
- [22] Green N, Rege M, Liu X, et al. Evolutionary spectral co-clustering[C]. International Joint Conference on Neural Networks. IJCNN, 2011:1074-1081.
- [23] Shi X, Fan W, Yu P S. Efficient Semi-supervised Spectral Co-clustering with Constraints[C]. International Conference on Data Mining. ICDM, 2011:1043-1048.
- [24] Huang S, Wang H, Li D, et al. Spectral co-clustering ensemble[J]. Knowledge-Based Systems, 2015, 84(C):46-55.
- [25] Kluger Y, Basri R, Chang J T, et al. Spectral biclustering of microarray data: coclustering genes and conditions.[J]. Genome Research, 2003, 13(4):703-716.
- [26] Cano C, Adarve L, López J, et al. Possibilistic approach for biclustering microarray data[J]. Computers in Biology & Medicine, 2007, 37(10):1426-1436.
- [27] Guan J, Qiu G, Xue X Y. Spectral Images and Features Co-Clustering with Application to Content-based Image Retrieval[C]. IEEE 7th Workshop on Multimedia Signal Processing, 2005.
- [28] Wieling M, Nerbonne J. Bipartite spectral graph partitioning to co-cluster varieties and sound correspondences in dialectology[C]. Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing. ACL-IJCNLP, 2009:14-22.
- [29] Xu G, Zong Y, Dolog P, et al. Co-clustering Analysis of Weblogs Using Bipartite Spectral Projection Approach[C]. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. Springer-Verlag, 2010:398-407.
- [30] Dhillon I, Mallela, S, Modha, D. Information-theoretic Co-clustering[C].ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2003:89-98.
- [31] Ghosh J. Bregman Co-clustering and Matrix Approximation A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation[J]. Journal of Machine Learning Research, 2007, 8(12):509-514.
- [32] Ye Y, Li X, Wu B, et al. Feature Weighting Information-Theoretic Co-Clustering for Document Clustering[C]. International Conference on Computer Science and ITS Applications, 2009:1-6.
- [33] Chiaravalloti A D, Greco G, Guzzo A, et al. An information-theoretic framework for high-order co-clustering of heterogeneous objects[C]. European Conference on Machine Learning. Springer-Verlag, 2006:598-605.
- [34] Huang Q, Chen X, Huang J Z, et al. Scalable Ensemble Information-Theoretic Co-clustering for Massive Data[C]. Proceedings of the International MultiConference of Engineers and Computer Scientists, 2012.
- [35] Cai R, Lu L, Cai L H. Unsupervised auditory scene categorization via key audio effects and information-theoretic co-clustering[C].IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP, 2005:1073-1076.
- [36] Wang P, Cai R, Yang S Q. Improving classification of video shots using information-theoretic

- co-clustering[C]. IEEE International Symposium on Circuits and Systems, 2005:964-967.
- [37] Ahmed M, Mahmood A N. Network Traffic Pattern Analysis Using Improved Information Theoretic Co-clustering Based Collective Anomaly Detection[M]. International Conference on Security and Privacy in Communication Networks. Springer International Publishing, 2014:29-41.
- [38] Castro P A D D, França F O D, Ferreira H M, et al. Applying Biclustering to Text Mining: An Immune-Inspired Approach[C]. International Conference on Artificial Immune Systems. Springer-Verlag, 2007:83-94.
- [39] Liang C, Leng Y. Collaborative filtering based on information-theoretic co-clustering[J]. International Journal of Systems Science, 2014, 45(3):589-597.
- [40] Wang J, Song H, Zhou X. A collaborative filtering recommendation algorithm based on biclustering[C]. IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems. CYBER, 2015:803-807.
- [41] Alqadah F, Reddy C K, Hu J, et al. Biclustering neighborhood-based collaborative filtering method for top-n, recommender systems[J]. Knowledge & Information Systems, 2015, 44(2):475-491.
- [42] 王德青, 朱建平, 王洁丹. 基于自适应权重的函数型数据聚类方法研究[J]. 数理统计与管理, 2015, 34(1):84-92.
- [43] 王德青, 朱建平, 刘晓葳, 等. 函数型数据聚类分析研究综述与展望[J]. 数理统计与管理, 2018, 37(1):51-63.
- [44] 林秀梅, 孙海波, 王丽敏. 智能信息处理的多指标面板数据聚类方法及其应用[J]. 数理统计与管理, 2016, 35(4):641-648.
- [45] Slimen Y B, Allio S, Jacques J. Model-Based Co-clustering for Functional Data[J]. Neurocomputing, 2018.